



SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi,
dan Teknik Informatika

<https://ejurnal.itats.ac.id/snestik> dan <https://snestik.itats.ac.id>



Informasi Pelaksanaan :

SNESTIK I - Surabaya, 26 Juni 2021

Ruang Seminar Gedung A, Kampus Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:

DOI : 10.31284/p.snestik.2021.1799

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Pembuatan Sistem Pencarian Hadis dengan Menggunakan Metode Pembobotan TF-IDF

Khisby Al Ghofari¹, Nanang Fakhrrur Rozi², Lukita Selmakaramy³, dan
Fernando Yoga Ariansyah⁴

Program Studi Teknik Informatika, Fakultas Teknik Elektro dan Teknologi Informasi,
Institut Teknologi Adhi Tama Surabaya^{1,2,3,4}
e-mail: khisby@gmail.com

ABSTRACT

In Indonesia, technology is developing very fast. One of the technological developments in Indonesia is the internet network which produces culture and the development of the virtual world. Unfortunately, there is also a negative part that very often occurs, namely misinformation or hoaxes. The hoax also targets the realm of religion, one of which is the misinterpreted hadith and false traditions. So that many people believe in and spread this hadith. As a result, many Indonesians have the wrong belief in the foundation of worship. One of the tools that make it easier to find information is a search engine. The search engine will generate the information we need with the keywords we enter. Unfortunately not all search engines can produce relevant search results. So that the information we are looking for can be mistaken. One method that can be used to increase the relevance of the search results is TF-IDF weighting. This study will implement a hadith search system using the TF-IDF weighting method which is useful for generating relevant searches for users. The search system produces a precision with average of 0.09 with the calculation of confusion matrix evaluation obtained from 6 keywords entered by 3 users.

Keywords: Hadith; Information retrieval system; Search engine; TF-IDF.

ABSTRAK

Di Indonesia, teknologi berkembang dengan sangat cepat. Salah satu perkembangan teknologi di Indonesia adalah jaringan internet yang menghasilkan budaya dan berkembangnya dunia maya. Sayangnya, hal tersebut juga mengandung bagian negatif yang sangat sering terjadi yaitu salah informasi atau *hoax*. *Hoax* tersebut

juga menasar ranah agama, salah satunya adalah hadis yang salah ditafsirkan atau hadis palsu sehingga banyak orang meyakini dan menyebarkan hadis tersebut. Akibatnya, banyak masyarakat Indonesia yang salah dalam meyakini pondasi ibadah. Salah satu alat yang memudahkan dalam mencari informasi adalah mesin pencari. Mesin pencari akan menghasilkan informasi yang dibutuhkan dengan kata kunci yang dimasukkan. Sayangnya, tidak semua mesin pencari dapat menghasilkan hasil pencarian yang relevan. Sehingga, informasi yang dicari dapat terjadi kekeliruan. Salah satu metode yang dapat digunakan untuk meningkatkan relevansi dari hasil pencarian tersebut adalah pembobotan TF-IDF. Penelitian ini akan mengimplementasikan sistem pencarian hadis dengan metode pembobotan TF-IDF yang berguna untuk menghasilkan pencarian yang relevan bagi pengguna. Sistem pencarian tersebut menghasilkan rata-rata *precision* sebesar 0,09 dengan perhitungan evaluasi *confusion matrix* yang didapat dari 6 kata kunci yang dimasukkan dengan 3 pengguna.

Kata kunci: Hadis; Mesin pencari; Sistem temu kembali informasi; TF-IDF.

PENDAHULUAN

Di Indonesia, teknologi berkembang dengan sangat cepat. Salah satu perkembangan teknologi di Indonesia adalah jaringan internet yang menghasilkan berkembangnya kebiasaan masyarakat Indonesia di dunia maya. Sayangnya, kondisi tersebut juga menghasilkan banyak hal negatif. Salah satunya adalah informasi yang sesat atau biasa disebut *hoax* [1]. Berbagai macam informasi yang salah di dunia maya juga menasar ke ranah agama dengan menafsirkan hadis secara berbeda atau pun terdapat hadis yang palsu.

Hadis palsu di Indonesia juga tersebar di masyarakat luas. Hadis tersebut dijadikan sebagai dasar ibadah. Sehingga beberapa informasi bisa jadi menyesatkan dan dapat digunakan oleh golongan tertentu untuk mencapai tujuan golongan tersebut [2]. Salah satu faktor berkembangnya hadis palsu adalah minimnya pengetahuan masyarakat tentang hadis. Para mubalig di mimbar-mimbar juga bisa menjadi peluang penyebaran hadis palsu yang bisa dipercaya oleh masyarakat [3].

Mesin pencari adalah alat bantu yang dimanfaatkan pengguna untuk mencari informasi secara cepat dan akurat dengan memasukkan kata kunci yang diinginkan. Sayangnya, tidak semua mesin pencari dapat memberikan hasil pencarian yang relevan dengan apa yang diinginkan pengguna. Salah satu metode yang dapat digunakan untuk menghasilkan informasi yang akurat dan relevan sesuai yang diinginkan pengguna adalah metode TF-IDF yang juga telah dilakukan pada penelitian sebelumnya, terbukti mampu menghasilkan informasi yang lebih baik. Dalam penelitian ini, Penulis akan mengimplementasikan metode TF-IDF dalam sistem pencarian hadis agar hadis yang ingin dicari tepat, akurat, dan tidak terjadi salah informasi [4].

METODE

Metode TF-IDF

TF-IDF merupakan suatu metode pembobotan kata terhadap suatu dokumen. Metode ini menggabungkan dua konsep perhitungan bobot yaitu frekuensi kemunculan kata dan *inverse* frekuensi dokumen yang mengandung kata tersebut [5]. Rumus perhitungan TF-IDF dapat dilihat pada Persamaan 1, 2, dan 3.

$$tf = 0,5 + 0,5 \times \frac{tf}{\max(tf)} \quad (1)$$

$$idf_t = \log\left(\frac{D}{df_t}\right) \quad (2)$$

$$W_{d,t} = tf_{d,t} \times idf_{d,t} \quad (3)$$

dengan tf adalah banyaknya kata yang dicari pada sebuah dokumen, $\max(tf)$ adalah jumlah kemunculan terbanyak kata pada dokumen yang sama, D adalah jumlah dokumen, df_i adalah jumlah dokumen yang mengandung kata t , idf adalah *inversed document frequency*, d adalah dokumen ke- d , t adalah kata ke- t dari kata kunci, dan W adalah bobot dokumen ke- d terhadap kata ke- t .

Diagram Alur Sistem

Proses pada sistem terbagi menjadi dua bagian. *Bagian pertama* berfungsi untuk pengumpulan data dengan metode *scraping*. Proses awal pengumpulan data dimulai dari (1) menyiapkan *browser*, (2) membuka *tools* yang digunakan untuk memprogram *bot* untuk *scraping* data menggunakan Google Colab, (3) menulis kode program, (4) menjalankan *scraping*, serta (5) menyimpan hasilnya pada file CSV. *Bagian kedua* adalah pembobotan kata. Pada bagian pembobotan kata, dimulai dari proses: (1) *load* data yang sudah disimpan pada file CSV, (2) dilakukan *pre-processing*, (3) disimpan lagi kedalam file CSV baru, (4) pembobotan kata menggunakan metode TF-IDF sesuai Persamaan 1, dan (5) menampilkan hasilnya kepada pengguna dengan kata kunci (*query*) yang dimasukkan.

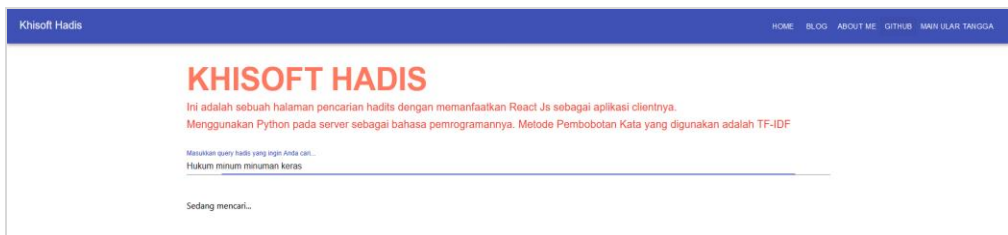
Dari rancangan sistem yang dibuat, kebutuhan perangkat keras dan perangkat lunak yang digunakan adalah komputer dengan prosesor Intel Core i7-8565U, RAM 8 GB, kartu grafis NVIDIA MX250, penyimpanan 512GB, serta sistem operasi yang digunakan adalah Windows 10 Home. Sementara, untuk kebutuhan perangkat lunak yang digunakan adalah aplikasi *text editor* Visual Studio Code, Google Colab Research, PyCharm, *browser* Mozilla Firefox, serta bahasa pemrograman yang digunakan adalah Python dan Javascript.

HASIL DAN PEMBAHASAN

Pada halaman implemetasi antarmuka pengguna (*user interface*) ini terdiri dari halaman awal, hasil pencarian, dan detail hadis.

Halaman Awal

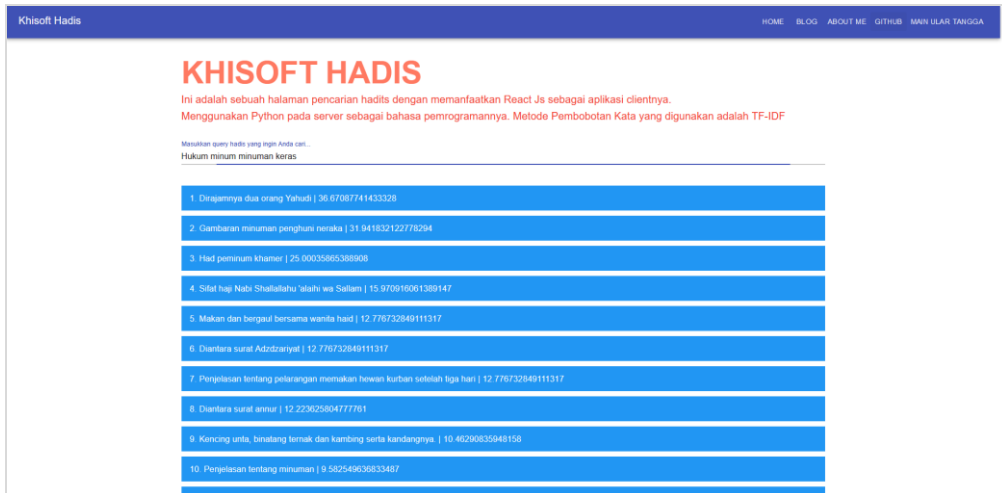
Pada halaman ini, pengguna dapat menuliskan kata kunci dari hadis yang dicari pada kolom pencarian. Ilustrasinya dapat dilihat pada Gambar 1.



Gambar 1. Halaman awal aplikasi.

Hasil Pencarian

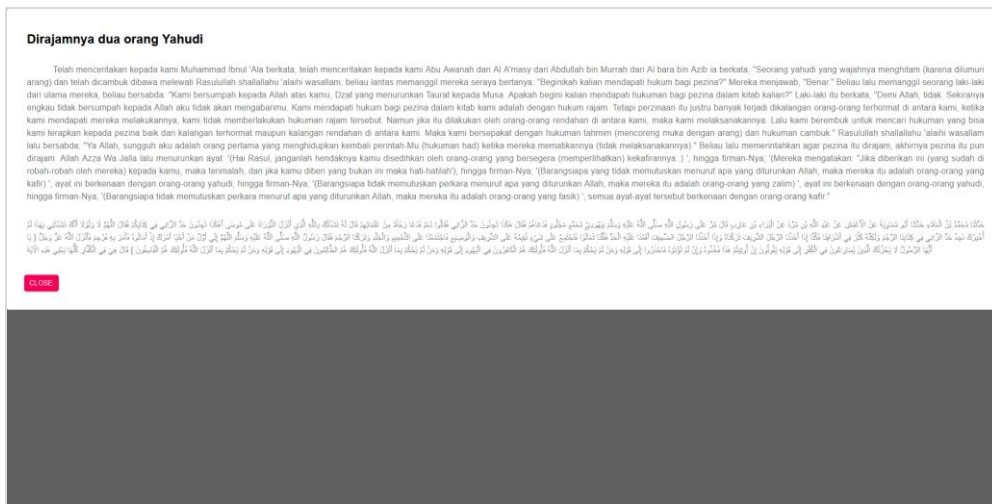
Pada halaman ini, pengguna akan mendapatkan hasil yang dicari dengan judul dan nilai bobot relevansi yang didapat. Pengguna dapat menekan pada kotak hasil pencarian untuk melihat penjelasan hadis. Ilustrasinya dapat dilihat pada Gambar 2.



Gambar 2. Halaman Hasil Pencarian.

Detail Hadis

Pada bagian ini, pengguna dapat melihat penjelasan dari hadits yang didapat. Ilustrasinya dapat dilihat pada Gambar 3.



Gambar 3. Halaman Detail Hadis.

Hasil Uji Coba

Hasil pengujian dengan *confusion matrix* dapat dilihat pada Tabel 1 dengan menggunakan 6 kata kunci (*query*) dan 3 pengguna. Kata-kata kunci yang digunakan yakni: (1) hukum minum minuman keras, (2) tidak sahur saat puasa, (3) tidak membayar zakat fitrah, (4) *lailatul qodar*, (5) puasa di hari raya, dan (6) pernikahan suami istri.

Tabel 1. Hasil pengujian.

Kata Kunci	TP	TN	FP	FN	Acc.	Precision	Recall
Pengujian Pengguna 1							
1	5	56,00	56,00	995,00	0,055	0,082	0,005
2	3	34,00	34,00	997,00	0,035	0,082	0,003
3	3	29,00	29,00	997,00	0,031	0,094	0,003
4	1	1,00	1,00	999,00	0,002	0,000	0,001
5	2	164,00	164,00	998,00	0,090	0,013	0,002
6	11	28,00	28,00	989,00	0,037	0,200	0,011
Pengujian Pengguna 2							
1	2	59,00	59,00	998,00	0,055	0,033	0,002
2	1	36,00	36,00	999,00	0,035	0,028	0,001
3	0	32,00	32,00	1.000,00	0,031	0,000	0,000
4	1	1,00	1,00	999,00	0,002	0,000	0,001
5	1	165,00	165,00	999,00	0,090	0,007	0,001
6	14	25,00	25,00	986,00	0,038	0,250	0,014
Pengujian Pengguna 3							
1	6	55,00	55,00	994,00	0,055	0,099	0,006
2	1	36,00	36,00	999,00	0,035	0,028	0,001
3	4	28,00	28,00	996,00	0,031	0,090	0,004
4	0	2,00	2,00	1.000,00	0,002	0,000	0,000
5	4	162,00	162,00	996,00	0,090	0,025	0,004
6	10	29,00	29,00	990,00	0,037	0,180	0,010
Rata-rata	4	52,33	52,33	996,17	0,090	0,090	0,000

KESIMPULAN

Dari hasil penelitian yang dilakukan, sistem pencarian tersebut menghasilkan precision average 0,09 dengan perhitungan evaluasi *confusion matrix* yang didapat dari 6 kata kunci yang dimasukkan oleh 3 pengguna. Hasil tersebut berpengaruh pada data yang didapatkan dan pengujian yang dilakukan. Saran untuk penelitian selanjutnya, sistem dapat ditambah dengan metode Cosine *similarity* dan dapat dibandingkan dengan sistem pencarian yang hanya menggunakan TF-IDF.

DAFTAR PUSTAKA

- [1] A. Majid, "Fenomena Penyebaran Hoax dan Literasi Bermedia Sosial Lembaga Mahasiswa Universitas Muslim Indonesia," *J. Komodifikasi*, vol. 8, pp. 228–239, 2019, [Online]. Available: <http://journal.uin-alauddin.ac.id/index.php/Komodifikasi/article/view/11329>.
- [2] Istianah, "Kritik Terhadap Penisbatan Riwayat Hadis: Studi atas Hadis-Hadis Palsu," *Riwayah J. Stud. Hadis*, vol. 4, no. 1, pp. 77–100, 2018.
- [3] A. Wahid, "Strategi Ulama Mengantisipasi Penyebaran Hadist Maudhu' di Kecamatan Peureulak," *Subst. J. Ilmu-Ilmu Ushuluddin*, vol. 20, no. 2, p. 119, 2018, doi: 10.22373/substantia.v20i2.5151.
- [4] A. Tirtana, A. Zulkarnain, and Y. D. Listio, "Pembuatan Sistem Pencarian Pekerjaan Menggunakan TF-IDF," *J. Ilm. Teknol. Inf. Asia*, vol. 13, no. 2, p. 91, 2019, doi: 10.32815/jitika.v13i2.389.

- [5] M. Nurjannah and I. Fitri Astuti, "PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING Mahasiswa S1 Program Studi Ilmu Komputer FMIPA Universitas Mulawarman Dosen Program Studi Ilmu Komputer FMIPA Universitas Mulawarman," *J. Inform. Mulawarman*, vol. 8, no. 3, pp. 110–113, 2013.